README: Stata and R codes

Credit Building or Credit Crumbling? A Credit Builder Loan's Effects on Consumer Behavior and Market Efficiency in the United States

Jeremy Burke, Julian Jamison, Dean Karlan, Kata Mihaly, Jonathan Zinman

DATA:

Data license agreements prevent us from sharing the data, but please contact dean.karlan@gmail.com if you have any questions about the code. For the sake of the transparency, we are sharing the code we used to generate the tables in the paper.

Do Files and R scripts:

Important: We used a combination of Stata and R codes to generate the tables for this paper. We used the "grf" package to run the causal forest model and Stata for the rest of the analyses. The "00 Run" do-file contains codes to run all the do-files and the R scripts. We tried to comment extensively on all the do-files to make it easy for the reader.

Note: Using the version of the "grf" package **(version 2.2.1)** available as of this writing (March 2nd, 2023)**,** we were unable to exactly replicate the results of the causal forest model (Table 4). The "grf" package used to produce the published results (version 0.10.4) is not accessible anymore. However, the differences are very small, statistically and economically speaking, and do not affect inferences (please compare the published vs. updated versions of Table 4 below). Other tables use the causal forest model results as inputs and their results also change in minor ways.

Table 4. Causal forest aggregate tests for CBL treatment effect heterogeneity (published version)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	1 = has FICO® Score 8			FICO® Score 8		
Endline:	6 mo	12 mo	18 mo	6 mo	12 mo	18 mo
Panel A: Aggregate test for treatment effect heterogeneity						
Mean forest prediction — Coefficient β_1 :	0.967	0.926	0.983	1.022	1.048	1.031
Mean forest prediction — std error :	(1.863)	(5.992)	(1.109)	(0.639)	(3.353)	(4.164)
Mean forest prediction — p-value ($\beta_1 = 1$) :	0.986	0.990	0.988	0.973	0.989	0.994
Differential forest prediction — Coefficient β_2 :	0.943	-1.057	-2.290	1.388	0.665	-0.263
Differential forest prediction — std error :	(0.690)	(1.027)	(0.959)	(0.492)	(0.523)	(0.876)
Differential forest prediction — p-value ($\beta_2 \leq 0$) :	0.086	0.848	0.991	0.002	0.102	0.618
Panel B: Average treatment effect by terciles of conditional average treatment effect						
Bottom tercile of CATE	-0.03	0.00	0.01	-9.98	-6.14	7.02
	(0.03)	(0.03)	(0.01)	(4.00)	(5.27)	(6.14)
Top tercile of CATE	0.01	-0.01	0.02	6.83	3.19	-5.50
	(0.02)	(0.03)	(0.04)	(3.93)	(4.36)	(5.07)
Difference of top tercile - bottom tercile	0.03	-0.01	0.01	16.80	9.33	-12.52
95% confidence interval range (+/-)	0.07	0.08	0.08	10.98	13.40	15.61
Number of observations	1413	1374	1330	1164	1126	1073

Unit of observation is a person-endline. For each column in this table-- each outcome-endline combination-- we ran a casual forest using the GRF package in R (Athey et al. 2019; R version 1.0.1, grf version 0.10.4) to predict the outcome listed in the column heading and obtain the CBL's conditional average treatment effects (CATE) on it. The p-values in Panel A show the probability that model is well-calibrated ($\beta 1 = 1$) and identifies homogeneous CATEs across observations ($\beta 2 \leq 0$). Panel B uses the predicted CATE for each observation to divide observations into CATE terciles (see Figure 3 and its discussion in the text for why terciles are warranted) and then estimates the OLS ITT separately for each tercile. The right hand side variables included in the causal forest for the binary outcome "1 = Has FICO® Score 8" are: age; number of adults in the household; number of

children in the household; standardized risk taking score; number of open trade lines; savings balance and combined savings and checking balance (both in hundreds of dollars, winsorized at 95th percentile); dummies equal to one if baseline survey is missing, credit report is missing, the participant is female, the participant's race is Black, the participant is married, the participant has attended college, the participant's household income is less than 30k, the participant is still an SLCCU member, and the participant has a non-CBL loan; and standardized indices of insecurity, self-control, attention to credit status, credit process knowledge, delinquency, new credit, and lack of liquidity. The right hand side variables included in the causal forest for the continuous outcome of FICO® Score 8 are those listed above, with the addition of baseline FICO® score and a standardized index of the amount that the respondent owes based on account balances. Sample sizes are lower here (than e.g., the number of individuals with data for each outcome in Table 3) because we are doing each outcome-endline combination separately, and because of missing values on input variables.

Table 4. Causal forest aggregate tests for CBL treatment effect heterogeneity (new version)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	1 = has FICO® Score 8			FICO® Score 8		
Endline:	6 mo	12 mo	18 mo	6 mo	12 mo	18 mo
Panel A: Aggregate test for treatment effect heterogeneity						
Coefficient β_1 :	1.005	0.950	1.012	1.007	1.026	1.063
Mean forest prediction std error :	(2.047)	(7.378)	(1.197)	(0.630)	(3.293)	(4.604)
p-value ($\beta_1 = 1$) :	0.998	0.995	0.992	0.991	0.994	0.989
Coefficient β_2 :	0.923	-1.088	-2.448	1.376	0.649	-0.164
Differential forest prediction std error :	(0.707)	(1.027)	(0.982)	(0.495)	(0.514)	(0.863)
p-value ($\beta_2 \leq 0$) :	0.096	0.855	0.994	0.003	0.103	0.575
Panel B: Average treatment effect by terciles of conditional average treatment effect						
Bottom tercile of CATE	-0.03	0.01	0.01	-9.09	-6.12	4.61
	(0.03)	(0.03)	(0.00)	(4.03)	(5.28)	(6.20)
Top tercile of CATE	0.00	-0.01	0.01	6.64	5.13	-3.12
	(0.02)	(0.03)	(0.04)	(3.91)	(4.33)	(5.09)
Difference of top tercile - bottom tercile	0.03	-0.01	0.00	15.73	11.25	-7.73
95% confidence interval range (+/-)	0.07	0.08	0.08	11.01	13.38	15.72
Number of observations	1413	1374	1330	1164	1126	1073

Note: Estimated using grf package [version 2.2.1] in R (version 4.2.1). See published version for addition notes.